

## REMARKS

Claims 1-19 are pending in the present Application. Claims 10-18 are withdrawn from consideration. Claim 4 has been canceled, and claims 1, 2, 3, 9, and 19 have been amended, leaving Claims 1-3, 5-9 and 19 for consideration upon entry of the present Amendment. No new matter has been introduced by these amendments.

Reconsideration and allowance of the claims are respectfully requested in view of the above amendments and the following remarks.

### Amendments to the Claims

Claims 1, 2, 3, 9 and 19 have been amended to better define the invention.

Support for the amendment to Claim 1 can be found in claim 4 as originally filed, on page 2 lines 13-17, and on page 4, lines 19-25.

Support for the amendment to Claim 2 can be found in the specification as originally filed in Table 1 (p. 5), rows 6 and 8, on page 4, lines 19-27, on page 6, lines 3-5, 14-18, 22-30, and line 34 to page 7 line 1, and on page 7 lines 6-9. Specifically, support for the phrase “a character representing a continuation of extracted differences” can be found in the specification in Table 1, row 8 and on page 6, lines 3-5, 17-18, 25-26, and line 34 to page 7 line 1, and page 7 lines 8-9. As originally filed, the specification discloses that the symbol “~” represents the continuation of a pattern (page 6, lines 3-4). This continuation of a pattern is explained further with reference to Fig. 3 as follows: “the “~1” represents that the number of the continued bases of the C pattern is one” (page 6, lines 17-18), “the “~6” represents that the number of the continued bases of the D pattern is six” (page 6, lines 26-26), and “[t]he “~3” represents that the number of the continued bases of the E pattern is three” (page 6, line 34 to page 7. line 1). The patterns referred to above are among the 6 possible patterns of the extracted difference between the reference sequence and the subject sequence (page 5, lines 4-23). Each of the example extracted difference patterns is converted into a string of characters that represent the extracted difference (p.6, line 1 – p. 7, line 12 ). Thus, the specification clearly teaches a character (~) representing a continuation of an extracted difference.

Claim 3 has been amended to state “a distance between the start position of the

extracted difference and the end position of the extracted difference” in order to clarify the claim. Support for this amendment is found at least in Table 1 (p. 5), row 6 and on p. 6, lines 7-8, 12-13, 20-21, and 28-29.

Support for the amendment to Claim 9 can be found on page 10, line 32 to page 11, line 2.

Support for the amendment to Claim 19 can be found in Claim 4 as originally filed, on page 4, lines 19-25 and on page 14, lines 17-19.

Claim Rejections Under 35 U.S.C. § 112, First Paragraph

Claims 2, 3, 9, and 19 stand rejected under 35 U.S.C. § 112, first paragraph, as containing subject matter which was not described in the Specification in such a way as to reasonably convey to one skilled in relevant art that the inventors, at the time the application was filed, had possession of the claimed invention. (Office Action (OA) 06/15/2007, page 2) Applicants respectfully traverse this rejection.

With regard to Claims 2 and 3, the Examiner stated that there does not appear to be adequate written support for the following phrases: “a number of base positions that characterize a feature of the extracted difference” (claim 2) and “whether a type of extracted difference occurs in succession in the subject sequence” (claims 2 and 3). (OA 06/15/2007, page 3) Claims 2 and 3 are now amended to more clearly define the invention, as described above.

With regard to Claim 9, the Examiner stated that there does not appear to be adequate written support for the phrase “two adjacent variations”. (OA 06/15/2007, page 3) Claim 9 is now amended to remove the phrase “two adjacent”.

With regard to Claim 19, the Examiner stated that there does not appear to be adequate written support for the phrase “is not a carrier wave”. (OA 06/15/2007, page 3) Claim 19 is now amended to recite specific positive limitations.

Reconsideration and withdrawal of the 35 U.S.C. § 112, first paragraph rejection of Claims 2, 3, 9 and 19 is respectfully requested.

Claim Rejections Under 35 U.S.C. § 112, Second Paragraph

Claims 1-9 and 19 stand rejected under 35 U.S.C. § 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

With regard to Claim 1 the Examiner has stated that the phrase “a character to represent the extracted difference” is confusing with respect to the “characters” recited in line 7, and that a similar issue is present in Claims 6 and 19. (OA 06/15/2007 page 4) Claims 1 and 19 are now amended to better define the invention. Specifically, Claim 1 (from which claim 6 depends) now states “a conversion code that corresponds to each character in the string of characters to represent the extracted difference”. More specifically, Claim 19 now states “encoding each character in the string of characters”.

Claim 4 is now canceled making all rejections directed to this claim moot.

Applicants believe the amendments overcome the issues. Applicants therefore request reconsideration and withdrawal of the rejection of claims 1-9 and 19 under 35 U.S.C. § 112, second paragraph.

#### Claim Rejections Under 35 U.S.C. § 102(b)

Claims 1, 6, 7, and 19 stand rejected under 35 U.S.C. § 102(b), as allegedly anticipated by Grumbach et al. (hereinafter “Grumbach”). (OA 06/15/2007, page 5) Applicants respectfully traverse this rejection.

Independent Claim 1 is directed to an apparatus for encoding a DNA sequence to achieve a high data compression ratio for storage or transfer which comprises: a comparative unit for aligning a reference sequence having known DNA information with a subject sequence to be compressed and extracting a difference between the reference sequence and the subject sequence; a conversion unit for converting the extracted difference between the reference sequence and the subject sequence into a string of characters and for outputting the string of characters; a code storage unit for storing a conversion code that corresponds to each character in the string of characters to represent the extracted difference; and an encoding unit for encoding the string of characters using the conversion code; wherein a type of the extracted difference comprises a start region mismatch between the reference sequence and the subject sequence; a blank representing there is no base in a base position in the subject

sequence corresponding to the reference sequence; a single base pair mismatch between the reference sequence and the subject sequence; a base insertion into the subject sequence; a multiple base pair mismatch between the reference sequence and the subject sequence, or an end region mismatch between the reference sequence and the subject reference.

Independent Claim 19 is directed to a computer readable medium having embodied thereon a computer program for a method for encoding a DNA sequence to achieve a high data compression ratio, the method comprising: aligning a reference sequence having known DNA information with a subject sequence to be encoded; extracting a difference between the reference sequence and the subject sequence; converting the extracted difference between the reference sequence and the subject sequence into a string of characters; wherein a type of the extracted difference comprises a start region mismatch between the reference sequence and the subject sequence; a blank representing there is no base in a base position in the subject sequence corresponding to the reference sequence; a single base pair mismatch between the reference sequence and the subject sequence; a base insertion into the subject sequence; a multiple base pair mismatch between the reference sequence and the subject sequence, or an end region mismatch between the reference sequence and the subject reference; and encoding each character in the string of characters using a conversion code corresponding to the character, wherein the computer readable medium comprises a ROM, a RAM, a CD-ROM, a magnetic tape, a floppy disk, or an optical storage medium.

To anticipate a claim, a reference must disclose each and every element of the claim. *Lewmar Marine v. Varient Inc.*, 3 U.S.P.Q.2d 1766 (Fed. Cir. 1987).

Grumbach teaches a compression algorithm for DNA sequences, which is based on the detection and encoding of factors and palindromes. (Page 876, paragraph 2) In addition, Grumbach teaches two modes for compression comprising a horizontal mode, where a sequence is compressed without reference to other sequences, and a vertical mode, where a DNA sequence A is compressed with respect to another sequence B. (Page 876, paragraph, 4) In this regard, Grumbach teaches that the factors used for the compression may be either an identical factor or a palindrome. (Page 881, paragraph 1) However, Grumbach does not teach all elements of the present invention.

While Grumbach teaches a compression algorithm for DNA based on the finding and encoding of identical factors or palindromes, it does not teach an algorithm for extracting a difference between the reference sequence and the subject sequence wherein the extracted difference comprises one of six possible patterns: a start region mismatch between the reference sequence and the subject sequence; a blank representing there is no base in a base position in the subject sequence corresponding to the reference sequence; a single base pair mismatch between the reference sequence and the subject sequence; a base insertion into the subject sequence; a multiple base pair mismatch between the reference sequence and the subject sequence, or an end region mismatch between the reference sequence and the subject reference. For at least this reason, Grumbach does not teach each and every element of independent Claims 1 and 19, and dependent claims 6 and 7. Since Grumbach does not teach or disclose all elements of the claimed invention it cannot anticipate the claims.

Therefore, Applicants request reconsideration and withdrawal of the § 102(b) rejection over Grumbach and, an allowance of the claims.

#### Claim Rejections Under 35 U.S.C. § 103(a)

Claims 1-7 and 19 stand rejected under 35 U.S.C. § 103(a), as allegedly unpatentable over Grumbach et al. in view of Robson et al. (hereinafter "Robson"). (OA 06/15/2007, page 12) Applicants respectfully traverse this rejection.

Establishing a *prima facie* case of obviousness requires that all elements of the invention be disclosed in the prior art. *In Re Wilson*, 165 U.S.P.Q. 494, 496 (C.C.P.A. 1970).

In making the rejection the Examiner stated that it would have been obvious to modify the apparatus of Grumbach by using the 4 bit code words and characters as taught by Robson et al. in order to perform searching in a more intelligent, structured and faster manner. (OA 06/15/2007, page 14) However, as discussed previously, Grumbach does not teach an algorithm wherein the extracted difference comprises a start region mismatch between the reference sequence and the subject sequence; a blank representing there is no base in a base position in the subject sequence corresponding to the reference sequence; a single base pair mismatch between the reference sequence and the subject sequence; a base insertion into the subject sequence; a multiple base pair mismatch between the reference sequence and the

subject sequence, or an end region mismatch between the reference sequence and the subject reference. Therefore, Robson must also teach this element missing from Grumbach in order for the obviousness rejection to be proper.

Robson teaches a natural sequence code for compression (title) for “the efficient storage and recovery of the gene and protein sequence data for the purpose of comparisons with a given sequence”, and that “[a]n example would be the search for a homologous sequence”. Robson further discloses that an example of another application would be the storage of sequences for the generation of molecular models. (Page 285, Col. 1, paragraph 2)

The Examiner stated that Robson describes using code to signify differences and their number, such as unknowns, blanks, and deletions. In this regard Robson teaches as follows:

The byte 01010000 indicates an unknown amino acid residue. The byte 0111000 indicates a ‘blank’ i.e. it is to be skipped and not used as part of the information used in making comparison. Two bytes representing two amino acids in 1 \* 5 bit code which are separated by 01110000 will be considered as contiguous and the ‘blank’ will not appear in the comparison.

(page 287, Col. 1, lines 1-8)

As discussed above, Robson teaches numeric amino acid (or nucleic acid) codes that are assigned based upon the amino acids (or nucleotides) in a given amino acid (nucleic acid) sequence and, discloses that these numeric sequences can be compared to determine homology between sequences or that they may be used for modeling studies. As such, the codes for blanks, unknowns, or deleted amino acids are assigned to the sequence itself; they are not assigned based on the extracted difference between a reference sequence and a subject sequence.

Robson, therefore, does not teach converting an extracted difference between the reference sequence and the subject sequence into a string of characters as required by independent claims 1 and 19. Further, Robson also does not teach that the extracted difference comprises a start region mismatch between the reference sequence and the subject sequence; a blank representing there is no base in a base position in the subject sequence

corresponding to the reference sequence; a single base pair mismatch between the reference sequence and the subject sequence; a base insertion into the subject sequence; a multiple base pair mismatch between the reference sequence and the subject sequence, or an end region mismatch between the reference sequence and the subject reference. For at least this reason, the combination of Robson and Grumbach does not teach or suggest each and every element of the claimed invention. Further, since Robson does not make up for the deficiencies of Grumbach, there would be no motivation to combine the references.

Applicants therefore believe that the Examiner has not made a *prima facie* case of obviousness over Grumbach in view of Robson. Applicants respectfully request a withdrawal of the obviousness rejection and an allowance of the claims.

Claims 1, 6-9, and 19 stand rejected under 35 U.S.C. § 103(a), as allegedly unpatentable over Grumbach in view of Selifonov et al. (US 2002/0183934 A1) (hereinafter "Selifonov"). Applicants respectfully traverse this rejection.

In making the rejection the Examiner stated that it would have been obvious to modify the apparatus of Grumbach by creating a variation sequence as taught by Selifonov. (OA 06/15/2007, page 16) Applicants traverse this rejection.

As discussed previously, Grumbach does not teach or suggest all elements of the claimed invention.

Selifonov teaches a method of generating libraries of biological polymers comprising generating a diverse population of character strings in a computer, where the character strings are generated by alteration of pre-existing character strings wherein the diverse population of character strings is then synthesized to comprises the library of biological polymers (nucleic acids, polypeptides, etc.). (Paragraph [0015])

Selifonov does not teach an algorithm to extract the difference between a subject sequence and a reference sequence wherein the extracted difference comprises a start region mismatch between the reference sequence and the subject sequence; a blank representing there is no base in a base position in the subject sequence corresponding to the reference sequence; a single base pair mismatch between the reference sequence and the subject sequence; a base insertion into the subject sequence; a multiple base pair mismatch between the reference

sequence and the subject sequence, or an end region mismatch between the reference sequence and the subject reference. For at least this reason, the combination of Selifonov and Grumbach does not teach or suggest each and every element of the claimed invention. Further, since Selifonov does not make up for the deficiencies of Grumbach, there would be no motivation to combine the references.

Applicants therefore believe that the Examiner has not made a *prima facie* case of obviousness over Grumbach in view of Selifonov. Applicants respectfully request a withdrawal of the obviousness rejection and an allowance of the claims.

It is believed that the foregoing amendments and remarks fully comply with the Office Action and that the claims herein should now be allowable to Applicants. Accordingly, reconsideration and allowance are requested.

If there are any additional charges with respect to this Amendment or otherwise, please charge them to Deposit Account No. 06-1130.

Respectfully submitted,

CANTOR COLBURN LLP

/Sandra L. Shaner/

By \_\_\_\_\_

Sandra L. Shaner  
Registration No. 47,934

Date: September 17, 2007  
CANTOR COLBURN LLP  
55 Griffin Road South  
Bloomfield, CT 06002  
Telephone (860) 286-2929  
Facsimile (860) 286-0115  
Customer No.: 23413